DOCUMENT RESUME

ED 026 763 EC 002 712

By-Amcoff, Sven
Intelligibility of the Speech of Deaf Children.
Uppsala Univ. (Sweden). Inst. of Education.
Pub Date 68
Note-40p.
EDRS Price MF-$0.25 HC-$2.10
Descriptors-Articulation (Speech), *Aurally Handicapped, *Communication (Thought Transfer), Deaf, Educational Experience, Elementary School Students, Evaluation Techniques, *Exceptional Child Research, Hard of Hearing, Hearing Loss, Listening Comprehension, Research Reviews (Publications), Speech Evaluation, *Speech Handicapped, Speech Skills, Speech Tests, Testing

To develop a simple, inexpensive technique to quantify speech comprehension of pupils (aged 7 to 13) in special schools for the deaf, the verbal responses to pictures by 111 pupils were judged for intelligibility by untrained listeners. Pupils were asked to identify 30 pictures; their taped replies were judged by listeners who wrote down what they thought they heard. The recording phase gave a measure of spoken vocabulary (SV) while the playback phase yielded a score for number of words correctly understood by the judges (UV). The ratio UV/SV gave a measure of speech comprehensibility (SC). Analysis of the data revealed that vocabulary and speech intelligibility increased with number of years in school (UV, p .001; SV and SC, p .005). Rank orders of speech quality were positively correlated with scales UV and SC; a moderate correlation existed between scales UV and SC and amount of residual hearing. The investigators concluded that judgment of speech intelligibility could be done reliably by untrained listeners, that quantity of speech production had little connection with speech comprehension within each grade, that results were similar for both understood vocabulary and speech comprehensibility, and that the magnitude of residual hearing is a poor predictor of both vocabulary and speech comprehension. (JB)

ED026763

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

report

The Institute of Education
UPPSALA UNIVERSITY
S:t Olofsgatan 12
75221 Uppsala SWEDEN
Head: W. Sjöstrand

38
1968

INTELLIGIBILITY OF THE SPEECH OF DEAF CHILDREN

by Sven Amcoff

EC 002 712

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY S. AMCOFF INSTITUTE OF EDUCATION, UPPSALA, SWEDEN TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

INTELLIGIBILITY OF THE SPEECH OF DEAF CHILDREN

Amcoff Sven. Intelligibility of the speech of Deaf Children. Rep. Inst. of Educ. No. 38 1968. A technique, consisting of allowing untrained listeners make judgements about the intelligibility of the speech of 111 deaf pupils, 7 - 13 years old, has been tested. These judgements were scored in two ways: firstly how many of the pictures the pupils could name (Scale SV) and then, how many of these that could be understood (Scale UV). A ratio between these UV/SV gave a third measure (Scale SC). The results show that judgement of comprehension was possible with high objectivity, that the level of homogeneity in the scale values was high, and that the constancy as regards distribution characteristics for different judges was high.

Introduction

The transfer of a message from a speaker to a listener can be roughly divided into 3 elements. The first is production of speech when sound is formed in the organs of speech. In this element are included articulation, intonation, etc. Next comes the acoustic link (sound waves) and then the auditive element.

A distortion of blocking of a message may occur in any of these elements. A speaker with a serious hearing-defect generally has heavy distortion of articulation which, even with other links intact, causes a deficient transfer of the message. The nature and level of this distortion is of interest for reasons of rehabilitation and for estimating the speaker's possibilities for communication.

Analysis of distortion can be done in any or all of the above-named

communication elements. The speech to be analysed can be determined either by the investigator's direction (e.g. by having the subject read a text out loud) or by use of spontaneous speech. The advantage with the former is that one knows what is going to be said. The disadvantages are that, for instance, the subject's pronunciation is influenced by the reading and that the method is useless for the illiterate. In the first case, the problem can be eliminated to a certain extent by making the subject memorise the text and speak from memory. The analysis of spontaneous speech is often very difficult when distortion is so great that what the subject has said or tried to say cannot be determined.

In this paper, <u>quality of speech</u> will refer to a general judgement of speech without direct concern with content; a listener's ability to reproduce the uttered message of a speaker will be called <u>speech intelligibility</u>.

1. <u>Anatomical and physiological analyses of speech defects</u>.

An attempt at objective description of speech deficiences has been made through analysis of the movements of the speech-organs, for instance by means of EMG where muscle-potentials are measured or by X-rays of the speech apparatus or parts thereof (Björk 1961, MacNeilage et al. 1967). Even other so-called palatographic methods have been used (Hadding-Koch & Pettersson, 1964) which makes it possible to study the physiological changes in the speech organs and even to relate them to the produced speech. Characteristic of these methods (except palatography) is the fact that they require enormous apparatus which, even with analysis of very short speech (a few seconds), yields a fantastic amount of data which is very time-consuming to analyse. Nor do these methods offer the possibility of measuring

objectively deviations from any "normal" movement and relating them to comprehension. Thus they can hardly be considered as practical methods for studies of individual differences in regard to quality of speech.

Another aspect of the articulation phase is the possibility for the listener to utilise the visual information from the speaker's facial and other visible motions. This means that if the listener does not see the speaker, a certain amount of information is lost. This information can be used (Steward 1967, O'Neill 1954) but there are no investigations as to whether deficient articulation reduces it, and if so, to what extent.

2. Acoustic analysis of speech.

One of the most important instruments for acoustic analysis of speech is the sound spectograph (Fant 1958, Potter et al. 1966). It yields a visual representation of speech waves with 3 parameters - time, frequency and intensity, Any speech can thus be objectively recorded and compared with any other similarly recorded. A quantitative description of such data is however difficult and, more important, timeconsuming. Besides, for the present, nothing is known about how comprehension of speech varies with quantitative deviations along these parameters from the "normal values". Thus current practice is to make subjective judgements as to whether deviations from the normal values may be assumed to influence a listener's comprehension of speech. Consequently, the method as such can hardly be used in practice for measuring individual differences in the quality of speech. The goal of spectrographic analysis is rather to gain information about points of phonetic nature (Fant 1958, 1962). In the not-too-distant future however, progress may be expected since resarch with synthetic speech

seems to make it possible to determine quantitatively the meaning of different speech-parameters for comprehension.

Other more simplified forms of apparatus exist which reflect individual parameters e.g. tonality, nasality, s-sound, etc. (Risberg 1966). There are used for instructional purposes in order to give pupils an immediate feedback to replace the missing or defective audition. The pupils must then judge their own quality of speech themselves. Martony (1967) and Phillips, et al. (1967) describe the construction of an apparatus which visualises intonation pattern and tonality; they have shown that this apparatus can be of good service in the instruction of these specific qualities.

Stark et al. (1967) have achieved good results using a more complex visualisation apparatus that reflects 1) amplitude contour 2) presence or absence of voicing and 3) spectral analys of speech sounds. Pronovost et al. (1967) have constructed an instrument that gives approximately round patterns with variations of roundness and bugle corresponding to different sounds. This apparatus proved useful for 1) reinforcing the correct production of [m, s, z, f, v, a, u, i and o] in words 2) reinforcing the presence of [p, t, k] and toneless fricatives [θ and ʃ] in words 3) showing that words consist of continuous flow of sound and 4) teaching pupils not to add "extra" phonems to words ending with a "stop". On the other hand, there were difficulties in patterning rhythm and duration.

These examples of visual aids in speech training have only been tried on a small scale and it is yet undetermined as to what extent pupils can use them for judging their own speech. Greatest success seems to be achieved when attention is concentrated on one parameter at a time eg. vocal pitch or s-sound. For such parameters, it should even be

possible to abtain objective measurement that can be related to perception. It would be thus possible to say, for instance, that at a certain scale value for the s-indicator, sound is perceived as [ch] and another as (s). There is a considerable step from this to a scaling of general speech quality.

3. Auditive analysis of speech.

The ear, together with the central nervous system, can analyse speech to an extent that, so far, cannot be imitated by any physical-technical apparatus. The technique for measuring speech quality or intelligibility through analysis can be divided into 3 groups:

a) Phonetically trained listeners transcribe what they understand, i.e. they analyse the flow of sound in individual sounds, make a judgement about intonation and transfer this to script. These reproductions can then be related to what the speaker has tried to say.

b) Laymen are asked to judge the quality of speech in different ways. These judgements are then converted into measures of speech quality with the aid of current psychological scale methods. c) Laymen are asked to reproduce (say or write) what they understand of a message and this is then related to what the speaker has tried to say.

a. Trained Judges

The classical project with this technique was done by Hudgins and Numbers (1942). In spite of its relative age, it is still the only inclusive investigation of speech deficiency in the deaf. The goal of the project was a) to identify errors of speech, categorise them and determine their occurence and b) to determine the relative effect of each error-category on speech comprehension. The study consisted of allowing 192 deaf children, between the ages of 8 to 19 years, to read 10 sentences, which were recorded on gramophone records.

These records were played back for a number of teachers-of-the-deaf (on an average 7) who were asked to write down what they understood. This was used as the basis of measurement of speech comprehension. Afterwards two experts listened to the records and transcribed the sentences phonetically. Rhythm was also similarly transcribed. Two main types of speech errors were found, namely an error in articulation (both consonants and vowels) and an error in rhythm. These categories were classified into a number of sub-categories. In the comparison between speech error and speech comprehension, it was found that rhythm and consonant error were of equal importance and that both rhythm and consonant error were more important than vowel errors.

Bjuggren (1954) has worked with the problem of estimating speech intelligibility in deaf children who are also illiterate. The children, 3-8 years old, were made to describe what they had seen on 34 pictures. The children's speech was recorded on tape, while at the same time, two experts wrote down all they could understand of what the children said. The children's kindergarten teacher was then allowed to listen to the tape several times in order to be able to interpret as many of the children's sounds as possible. A speech comprehension ratio was extracted from this material in two ways. In the first, any child having more than 10 % completely unintelligible (of his total vocabulary) was excluded. Then, for the others, a ratio between completely understood sentences and interpreted sentences was constructed. In the second case, all the children were included and the ratio was made between perfect sentences and the total number of sentences.

Both methods showed that intelligibility was a direct function of vocabulary i.e. that the proportion of interpreted words increases with an increase in vocabulary. Vocabulary was also directly related to the length of sentences. Two cases deviating from this pattern had been exposed to intensive auditory training which had resulted in an increased vocabulary and longer sentences, but not in increased speech intelligibility. Those subjects who had a hearing-loss in the high frequency speech area and, with small rests in the bass register, did not achieve 30 % intelligibility. The two above-mentioned cases belonged here. The author points out however that a number of other factors may influence speech intelligibility but that the limited size of the material does not permit further interpretation.

Steward (1967) describes an investigation - still not completely analysed - in which the goal was to establish phonetic range of the speech of deaf children. Data were collected from 300 profoundly and severaly deaf children who were between 3 to 15 years old; these data consisted partly of tests of articulation (individual words) and partly of spontaneous speech. No technical aids were used, excent for a tape-recorder, and the speech sounds have been described in articulation-phonetic terms. So as to facilitate international comparisons, the IPA- alphabet has been used as the basis for transcription. In addition furthersigns have been placed in order to describe some aspects of the lip read pattern (i.e. the placement of the lips and the tongue, etc.).Steward also points out a number of difficulties in the analysis of only hearing. He mentions as an example the fact that the judges were often in disagreement as regards rhythm and length. The final product of the analyses will be presented in the form of a very careful phonetic transcription.

b) The judgement of speech intelligibility by laymen

In a noteworthy Swedish investigation by Hansson (1905) it was stated that intelligibility of the speech of children with serious hearing defects was very limited. The investigation, which was divided into two parts, aimed 1) at grading speech intelligibility in Grade 8 pupils at schools for the deaf and 2) at studying the connection between speech comprehension on the one hand and, on the other hand, level and type of hearing loss.

The subjects consisted of 22 pupils from a Deaf school's finishing class (not the Birgitta school)[1] for point 1 given above and, for point 2, five pupils from the finishing classes of all the Deaf schools. The stimulus material consisted of 9 vowels, 18 words with the commonest consonants, both initial and final, and 4 ordinary sentences (5-8 words) with a minimum of redundance.

The stimuli were printed in block capitals on cards which were then shown to the pupils who were instructed to learn what they read. Afterwards the cards were removed and the pupils asked to repeat what they had read on the cards. This was recorded on tape. The goal of the procedure was to tape speech which was, to the greatest possible extent, spontaneous. It is not clear whether or not the pupils had hearing-aids during the experiment. Individuals, without previous experience with defective speech, were selected as judges. Each taped speaker was judged by two persons who were asked to write down what they thought they heard. For each new speaker, one of the judges was changed. For point 1 above, it was shown that 27 % of the words, 9-6 % of words with short vowels and 7.1 % of words with long words were understandable to the listeners. The corresponding precentage for sentences was 18.2.

1) See page 12.

Of the 22 speakers, 10 had not produced any intelligible sentence, while one pupil achieved 62.5 % intelligible sentences. As regards point 2, a definite connection was shown to exist between type and level of hearing-loss and intelligibility. The 15 pupils were divided into three groups depending on the degree of hearing-loss. Group A was classified as having a loss greater than 100 db at 500 Hz, group B had relatively good hearing in the bass register but a hearing curve that diminished to around 100 db at 2000 Hz. Group C had less of a loss than the previous group and which loss was relatively evenly--distributed over the entire frequency-area.

Group C had results that differed markedly from the others and had perfect intelligibility when the stimulus material consisted of sentences. The percent correct in group B was 27.5, and in group A 12.5. Background factors other than hearing may contribute to these differences and it would have been interesting to know whether the groups differed with regard to language level in general, ability, method of instruction, etc.

c) Psychological Scaling Techniques

Psychological scaling techniques have been used in order to obtain a quantitative measure of articulation errors or specific information about special aspects of speech, e.g. nasality and tonality.

Sherman and Moodie (1967) investigated four different scaling techniques with respect to their use for grading articulation errors in short segments of fluent speech. They compared the following methods: equal appearing intervals, successive intervals pair comparison case V, and constant sums, The stimuli consisted of 5-seconds segments from fluent speech of children between the ages of 5 and 10. The segment represented speech varying from normal to considerable defective. The judges consisted of 350 Psychology students divided

into groups of 100 to 125, who judged according to one or two of the chosen methods.

The methods were evalated according to correlation and level of linearity between the scales and the level of homogenity within each scaling technique. Equal appearing intervals was shown to be the best, since it was reliable, required relatively simple calculation and agreed well with the homogeneity achieved with the method of successive intervals. Pair comparison, case V,showed a definite lack of homogeneity which, the authors thought, depended on variation in discriminal dispersion. Constant sums, gave a completely different resultat from the other methods. The reason for this could not be ascertained.

In an experiment by Prather (1960), comparisons were made between different forms of direct magnitude estimation and also between these measurements and the scales used in Sherman & Moodie's investigation. The material and judges were essentially similar. Prather's results demonstrated high agreement among the different scales, and, in opposition to Sherman and Moodie, found that even constant sums yielded the same results as the other scales.

In other words, results indicate that psychological scaling methods that give equidistant measures and even sometimes measures with an absolute zero point could be used for grading individual differences in speech quality. However, the literature records no attempts to analyse further differential psychological data achieved by such techniques.

From the aspect of differential psychology, the most important questions are (1) whether different segments from one and the same subject's spontaneous speech are given the same scores i.e. whether the

individual scale values are reliable estimations of scores on a more general dimension and (2) whether the individual scores are systematically related to some other differential psychological variable. As suggested above, these questions have not been answered with the aid of psychological scaling methods.

4. Purpose and Method

The goal of the project was to develope a simple and not-too-costly technique through which speech comprehension, for pupils in the age range 7-12 years in special schools for the deaf,could be quantified. These pupils have especially in the low levels a weak language development and poor reading ability. Thus, we chose the use of pictures of simple everyday objects as the stimulus material for the pupils to interpret.

We assume that there is a high correlation between speech comprehension determined through words and that determined through sentences, even though rhythm occurs rather limitedly in individual words. Calculating the rank correlation between words (group 1) and sentences in the previously-mentioned experiment by Hansson (1965) gives a vaule of rho =.59 (N=22). This indicates that a positive relation exists but since the tests used differentiated poorly, no definite estimation of the magnitude of the correlation can be made.

The pupils' speech was recorded on tape which reduces intelligibility since lipreading is rendered impossible and also both recording and play-back may alter the acoustic qualities of the speech. These factors, however, are assumed to have the same effect on all speakers and should thereforenot interfere with the relative rank-order between them. The judges were students with normal hearing and with no previous knowledge of speech deficiencies. They were given the task of writing down the words they thought they heard. The project was divi-

ded into two stages. In the first, which was a pre-run, the participants were pupils from the Birgitta school in Örebro. The reason for this was that if these pupils who have, on average, better hearing than pupils at other special Deaf schools, did not understand the instructions and procedure, and if the listeners could not judge their speech, then the method would not be suitable for pupils with more severe hearing-defects.

From the point of view of the method, however, the test-run gave such favourable results that the test was repeated with the same procedure and with pupils from Grades 2, 4 and 6 from two other special schools for the deaf. Data from both stages have been used for the evaluation of the technique. However, no strict comparison can be made between the Birgitta school results and those of the two other schools, since the tests were done at different times of the school year.

a) Sample

The sample consisted of 118 pupils from Grades 2,4 and 6 from the three special schools for the deaf. A detailed distribution is given on the next page.

		Sample	Lost cases	Remainder
Kristina school School 1	Grade 2	7	0	7
	" 4	16	0	16
	" 6	6	0	6
Manilla school School 3	Grade 2	16	1	15
	" 4	10	0	10
	" 6	18	2	16
Birgitta school School 5	Grade 2	18	3	15
	" 4	27	1	26
All schools	Grade 2	41	4	37
	" 4	53	1	52
	" 6	24	2	22
Total		118	7	111

The lost cases were pupils who were absent on the day the recording was made and were considered as completely random.

b) Apparatus

The recordings were made with Tandberg and Revox tape-recording, using a tape-speed of 9.5 cm per second (3 3/4 ips), and with a Tandberg T M 3 microphone.

c) Procedure

The recording was done in two sessions. At the end of the fall term of 1966, speech from School 5 pupils was recorded and at the end of the spring term 1967, speech from the other pupils was recorded. Data from School 5 were regarded as a pre-run, but since the method was exactly the same as for the other pupils, these data were included for judgement of the method. The procedures can be considered as an extension of the pre-run.

The recordings were made with one pupil at a time. The experimenter was introduced to the whole class by the teacher. Recordings were done in the school's acoustics room or, where no such room existed, in another undisturbed area. The pupils were seated at a table with the microphone about 30 cm away. The experimenter sat directly opposite on the other side of the table. The stimuli consisted of 33 pictures of ordinary objects. There of the pictures were given as practice trials, and were presented before the test.

The practice trials consisted of showing one picture at a time with the experimenter asking "What is it?" If the pupil understood immediately what was required, the recording of the test series began. Otherwise, the other practice pictures were used until the pupil had completely understood what he should do. Keeping in mind the later judgement of the recording, the 30 remaining pictures were arranged randomly into five different series. First, the name of the child was recorded and then the random series number, following which all the pictures were shown one by one. For each picture, the question "what is it?" was repeated. If the pupil shook his head or said "Don't know" this was repeated by the experimenter to ensure its being recorded. During the recordings, all pupils used their hearing-aids.

d) Play-back

The judges consisted of 4 (in the pre-run there were two more) 20-30-year-olds with normal hearing. These judges worked in twos in a quiet and undisturbed milieu but judged independently. Each listener was provided with an answer-sheet for each pupil (appendix 5); instructions were then read out loud (appendix 6) and questions, if any, were answered. The judges were allowed to work at their own speed, which meant that they could take a break whenever they wanted. Repetitive play-backs were not allowed. The tapes were randomised for

judge-pairs so as to eliminate possible practice effects. The listeners were asked to write down what they thought they heard and also indicate whether they were certain of their answers or not.

e) Tabulation of the judges´responses

Version 1

The responses were tabulated into three categories: correct, wrong and don´t know. The classification "correct" was made if the written word corresponded to the picture. Some pictures proved to be ambiguous. If three out of four judges understood the same thing, such an answer was classed as "correct". All other words and non-responses were classified as wrong. "Don´t know"was reserved for the case where the pupils themselves could not give an answer.

Version 2

The responses were classified as above with the variation that only responses deemed correct for sure were used.

5. Statistical Analysis

a) The usefulness of the method

From the aspect of methodology, the main questions to be investigated were (1) the degree of objectivity in the judgements, i.e. to what extent the judges rated the individuals in the same order (2) the degree of homogeneity in the scale values i.e. to what extent the tasks of the test graded the individuals in the same way and (3) the degree of constancy with regard to distribution characteristics for different judges.

b) The relation between judges

Since the same judges were used for the pre-run, these data were included in the calculations. The total number of pupils was 111. The correlation calculations were thus based on 111 pupils for four judges.

Table 1. Inter-correlation between judges. Version 2 is placed above the diagonal and Version 1 below it. N = 111 for judges 1-4.

		Judge			
		1	2	3	4
Judge	1		.921	.912	.913
	2	.959		.915	.913
	3	.967	.961		.903
	4	.955	.946	.972	

As can be seen from Table 1, agreement between the judges is high. It can be seen that when judgements are made regardless of certainty, the relationship increases. The high correlation may be regarded as an expression of high objectivity in judgement. Homogeneity was also calculated for each judge for both versions. It can be seen from Table 2 that homogeneity, calculated according to Kuder-Richardsson's formula 20, is high and that variation between judges is very small. This means that all the items of the test measure practically the same dimension or dimensions and that the choice of words did not have any significant importance. With regard to the two ways of classification, Version 1 was shown to give somewhat higher scores but the differences are small. Since even the correlation between versions among the judges is high, (Table 3), a combination of Version 1 and 2 would not give any increased precision and therefore, from the aspect of measurement, it is unnessesary to employ this rather cumbersome technique.

Table 2. Homogeneity coefficients according to Kuder-Richardsson's Formula 20 for 4 judges for Version 1 and 2. N=111, no of items =30.

	Judge 1	2	3	4
Version 1	.92	.93	.92	.91
Version 2	.93	.92	.89	.90

Table 3. Product Moment Correaltion between Version 1 and 2 for each judge. N=111.

	Judge 1	2	3	4
Ver. 1 - ver. 2	.94	.95	.83	.84

Table 4. Means and Standard Deviations for each judge and version

		1	2	3	4
Ver. 1	m	9.59	10.12	9.96	9.65
	s	7.41	7.87	7.66	7.23
Ver. 2	m	7.75	6.89	4.91	4.58
	s	7.38	6.96	5.41	5.45

The means and standard deviations for each judge and for both versions are presented in Table 4. The judges have the same means and the same variation in Version 1. Analysis of variance among them gives $F\ 3\ ;\ 436 = <\ 1$.

In version 2, there are considerable differences between the judges and an analys of variance of the differences between the judges give $F\ 3;436=6.41\ p < .001$. Thus the judges use different norms for their decisions of certainty. It is therefor unsuitable to use data from this version. Further analysis is thus based only on Version 1.

c) The choice of scales within Version 1

The test procedure may be divided into two aspects - recording and play-back - each of which constitutes a variable. From the first aspect, recording, a measure of the pupil's spoken vocabulary can be obtained, i.e. for how many pictures the pupil had a spoken expression. Hereafter, this will be referred to as spoken vocabulary (SV). From the second aspect, judging on play-back, we can obtain how many of the spoken words are understandable and correct, and this is the variable that was analysed in the previous section. Hereafter, this will be referred to as understood vocabulary (UV) and constitutes the average number of words understood by the 4 judges.

This variable may be taken as a measure of speech intelligibility, but since SV is different for different pupils, it would be plausible to correct for this difference. This may be done, for instance, by calculating UV only from those items which are common to all the pupils' SV. A certain amount of information would, however, be lost since these items would be determined by the pupil who has the least SV.

Another way would be to use the ratio of UV to SV (UV/SV). This ratio then expresses speech comprehension as that part of SV which is understandable. In order to facilitate the calculation, the ratio was expressed as a percentage, i.e. Speech comprehensibility (SC) is defined as, $SC=UV/SV \cdot 100$.

This measurement has the disadvantages that it is based on different pupils and that the correlation becomes greater as SV becomes smaller.

The scale UV has a very high precision since it is based on four judgements, each of which having a high precision. SV is constant for all judges since the experimenter has already determined whether the pupil

has given an answer or not. The ratio between these scales, i.e. SC, must also therefore be assumed to have satisfactory precision. From this point of view, it therefor does not matter which scale is chosen, UV or SC. The use of scales UV or SC as measures of speech intelligibility has each its own disadvantages and it is therefor of interest to study their relationschip to each other and to other variables.

d) Distribution, Md, means and deviations

For each of the three scales - UV, SV and SC - and for each grade a frequency polygon(append. 2-4) has been drawn. Medians, means and deviations have been calculated for these distributions (Table 5). The distributions for the variables tend to be skewed or, in certain cases, bimodal. Statistical analysis has thus been done with non-parametric methods.

Table 5. Medians, means and deviations for the variables UV SV and SC for every grade. N_2=37, N_4=52, N_6=22.

		åk 2	åk 4	åk 6
	Md	6.00	10.50	15.00
UV	M	6.59	10.37	14.32
	S	4.98	8.10	6.09
	Md	23.00	25.94	29.50
SV	M	22.73	25.25	27.63
	S	6.12	3.90	3.88
	Md	29.25	47.00	52.5
SC	M	30.49	39.60	50.55
	S	22.41	28.86	20.50

e) Differences between grades.

In order to determine whether the scales differentiate between grades, the differences have been tested for significanse with the median test. (Appen. 3; Table 6) (Guilford 1965, p.258).

Table 6. Medain Test between Grades 2.4 and 6 for Scales UV, SV and SC.

	chi^2	df	p
UV	14.65	2	p <.001
SV	10.74	2	p <.005
SC	12.47	2	p <.005

The differences between grades are significant for all scales which mean that both vocabulary and speech intelligibility increase with the number of years in school.

f) Relationship between the scales.

The relationship between the scales within each grade has been estimated by calculating the rank correlation (rho) between them (Table 7) (Guilford 1965, p. 305).

Table 7. Rank correlation (rho) between UV, SV and SC within each grade.

	åk		åk 4		åk 6	
	SV	SC	SV	SC	SV	SC
UV	.076	$.915^{xxx}$	$.472^{x}$	$.981^{xxx}$	.415	$.974^{xxx}$
SV		-.256		$.347^{x}$		.250

x = p <.05 xx = p <.01 xxx = p <.001. The level of significance throughout for p <.05.

The relationship between UV and SC is very high and introducing SC gives very little further information beyond what already exists in UV (Table 7). In grade 4, there is a moderate but significant correlation between even SV and the other scales. Thus the scales UV and SC share a large common variation while SV, for the most part, measures something else.

Vocabulary as measured by SV is thus relatively independant of speech intelligibility. Both of the measures for speech intelligibility are practically interchangeable.

g) Item Analysis

Since homogeneity is high for each judge and each grade (Table 2) any detailed item analysis was deemed unnecessary (Scale UV). The average proportion of correct responses (P) for judges 1-4 have been calculated for each grade (appen.1). The means (p) in Grades 2, 4 and 6 are .22, .35 and .48 respectively. There is thus a gradual increase in p from Grade 2 to Grade 6. Since the mean p in Grade 6 is only .48, the test seems to be useful even for higher grades. P for individual items all increase from Grade 2 to Grade 4, but Grade 4 to Grade 6 there is no increase for 3 items.

Biserial correlations have been calculated within each judge and grade, between each item and total score on the remaining items. All the relationship are positive expect for one item in Grade 6 (appen. 1). From this, it is clear that any change in item composition can hardly be considered necessary. In that event, it would consist of introducing easier tasks by having the pupils say individual sounds, (e.g. vowels) or complete sentences. However, the entire test procedure would probably then become considerably more complicated.

h) Correlation with speech quality.

The pupils from School 5 have been rankordered within each grade by a teacher for quality of speech. The same was done with pupils from School 3 but there the rank-ordering was done by 4 recently graduated teachers who listened to archive tapes of speech from the said pupils.

Table 6. Rank correlation (rho) between estimations of speech quality and the scales UV, SV and SC. + = p < .5

		åk 2		åk 4		åk 6	
		n	rho	n	rho	n	rho
School 3	UV	14	.15	10	.62+	10	.78+
	SV		-.31		.06		.16
	SC		.23		.87+		.76+
School 5	UV	15	.43	26	.77+		
	SV		-.05		.41		
	SC		.49		.75+		

In comparing these rank-orders with the constructed scales from the experiment, a clear tendency is seen (Table 6). The scales UV and SC have a similar correlation with the rank-orders of speech quality in both the schools. The rank-correaltions (rho) are low for Grade 2 but high for Grade 4 (even for Grade 6 in School 3). Thus, ranking according to speech quality and judgement of speech intelligibility seem to be based, to a large extent, on the same characteristics for Grade 4 and 6, but not for Grade 2.

Vocabulary (SV) does not seem to have influenced rank-order according to speech quality since the correlation throughout is low or negligible. No difference between the scales UV and SC in relation to speech quality can be observed.

i) Correlation with hearing loss

Audiograms have been collected for 34 pupils from Grade 2, 51 from Grade 4 and 22 from Grade 6. This means a loss of 4 pupils distributed as follows: 3 in Grade 2 and 1 in Grade 4. All were from School 5. The average hearing-loss in db for 500, 1000 and 2000 Hz varied somewhat between the grades and was greatest in Grade 2. The largest variation was in Grade 4 which also contained the largest number of pupils (Table 9).

Table 9. Medians, means and variations for average-loss in db for 500, 1000 and 2000 Hz distributed according to Grades. Max. Loss=130 db.

	Md	M	S	N
Grade 2	103.25	101.56	15.04	34
Grade 4	98.00	97.75	16.03	51
Grade 6	93.50	97.95	12.33	22

In comparing hearing-loss with the variables UV, SV and SC, the phi-coefficients have been calculated. Dichotomy has been done along the median. Significance testing was done by transforming phi to the corresponding chi^2 -value with one degree of freedom. (Guilford, 1965, p. 335-5).

Table 10. Phi-coefficients for the relationship between average hearing-loss in db for 500, 1000, and 2000 Hz and the variables UV, SV and SC, along with corresponding values of x^2

	Grade 2			Grade 4			Grade 6		
	phi	chi^2	p	phi	chi^2	p	phi	chi^2	p
UV	.41	5.77	p <.02	.35	6.07	p <.02	.09	1.83	p >.05
SV	.18	1.0	p >.05	.02	< 1	p >.05	-.09	1.83	p >.05
SC	.29	2.94	p >.05	.43	12.25	p <.001	.09	1.83	p >.05

Residual hearing has a significant relation with the variable UV in Grades 2 and 4 and with SC in Grade 4 (Table 10). The variable SV has no relationship with hearing-loss in any grade.

The computed correlation between the scales and hearing-loss are moderate. It seems as though the magnitude of residual hearing, as these are considered to be, is a poor predictor of both vocabulary and speech comprehension. In calculating average hearing-loss, possible residual hearing over 2000 Hz was disregarded. Distribution of the audiograms was therefor done according to the shape of the hearing curve (Hansson , 1965, 1967, p. 131).

Group 1. Pupils with a relatively homogeneous hearing-loss over the entire frequency area -"flat loss".

Group 2. Pupils with relatively good residual hearing in the lower frequencies but with a hearing-loss that drops to about 100 db at 2000 Hz.

Group 3. Pupils with a hearing-loss greater than 100 db at 500 Hz.

Grouping could not be done without a certain amount of arbitrariness. This was especially so in deciding between Group 2 and 3. The majority of pupils ended up in Group 2 and 3. In grade 4 however, there were 9 pupils in Group 1 (Table 12) but in Grade 2, there were only 2 and in

Grade 6, 3 pupils in this group. The correlation was still calculated between the groups and the scale UV for all grades and tested for significance with chi^2-test. The scale UV was dichotomised at the median.

Table 11. Chi^2-test of the correlation between UV and grouping according to the hearing curve. Fg = 2.

	Grade 2	Grade 4	Grade 6
chi^2	5.26	12.27^{xx}	3.62

xx = p <.01

As can be seen in Table 11, a significant correlation was found only in Grade 4. The size of the correlation has been estimated to be .44 and is composed of the contingency coefficient corresponding to chi^2. The corresponding contingency coefficient for scale SC is .38. The correlation between UV-SC and hearing-loss can be regarded as moderate to low.

Table 12. Contingency table for calculating the correlation between UV and grouping according to the hearing-curve in Grade 4.

		Type of hearing-curve 1	2	3		
UV	>Md	0	11	5	25	C = .44
	<Md		14	12	26	

On comparing Tables 10 and 11 agreement is found for Grades 4 and 6 but not for Grade 2, such that a significant relationship is found between UV and average hearing-loss but not between UV and grouping according to the hearing curve. This may depend on the low number of pupils in Group 1 or on the fact that the division between Groups 2 and 3 was

all to arbitrary. It is also possible that the differences in conditions between Groups 2 and 3 are not so great and that distribution according to average hearing-loss is a more sensitive measure within these groups.

It is also possible that other factors of importance for speech development e.g. ability, other handicaps, time of injury, etc. are not evenly distributed among the groups and thus contribute towards lowering or eliminating the relationship. It may be seen from the results of Grade 4, however, that an individual with hearing-loss belonging to Group 1 - "flat loss" - has greater possibilities of acquiring comprehensible speech. All 3 pupils who belonged to this group lie above the median in the scale UV (Table 12).

6. Discussion

Analysis of the data has shown that judgement of speech intelligibility can be done in a reliable way by untrained listeners. It has also been shown that the quantity of speech production has little connection with speech comprehension within each grade. But even if the correlation for all grades is calculated, (UV-SV), phi becomes .35 (chi^2=13.60,p <.001) which is surprisingly low bearing in mind that both variables differentiate between grades.

As regards introducing SC as a ratio between UV and SV, one would expect from the above a variable with characteristics other than those of either UV or SV. It was found, however, that UV and SC gave practically the same results. This probably depends on the facts that distribution in SV is strongly negatively skewed, that the variation in UV is greater than in SV, and that the correlation between UV and SV is low. Thus, one cannot obtain any more information from UV by placing it in relation to SV. UV can therefore just as well be chosen

as SC as a measure of speech intelligibility and has many advantages from the aspect of calculations.

The validity of scale UV is troublesome since there is no consistent criterion for speech intelligibility. The positive correlation, found between UV and teacher-estimations of speech quality, still give a certain validity since it may be assumed that good speech quality also gives good speech intelligibility. A direct measure of speech intelligibility would be to let the pupils try to convey a message in a practical situation and grade them according to how well they succeeded. It is probable that such a "face-to-face" situation would give better results than if the message was transferred via tape, for example; but it is also probable that the rank-order among speakers would be very similar in both situations.

With the same kind of reasoning, it may be deduced that a test with only words or sounds would give the same order among the speakers. It would nonetheless be intresting to investigate how these relationships would look.

The correlations between scales and hearing-loss look all alike, regardless of whether the average hearing-losses for 500, 1000 and 2000 Hz are calculated or whether classification is made on the basis of hearing-curve. The correlations are significant between UV-SC and residual hearing when deviation is sufficiently large in the group. The magnitude of the correlation is, however, moderate since audiograms with sinus tones as stimuli detect only a small part of the hearing ability.

Valuable supplementary measures of hearing could be obtained through tests that measure the ability to discriminate between frequencies and between phonem or phonem compositions within words e.g. in the form of a so-called rhyme test (Risberg 1967). It would be desirable to perform such tests and relate them to the development of

speech, since such information could form the basis both for developing an educational program and for prognosis of the pupil's ability to make use of training in hearing and speech.

For investigations of the type presented above in brief, as well as for evaluation of certain aspects of instruction, the procedural technique developed in this investigation appears to be useful. The distribions of the UV-scale do not seem to prevent a transformation to a stanine scale which would facilitate its use.

It is obvious that children of normal hearing (down to age 4-5 years old) should have achieved perfect or nearly perfect speech intelligibility. No such comparison, however, has been done. The discrepancy between normal hearing and severe hearing-loss with regard to speech intelligibility is most likely exceedingly large.

Acknowledgements

This project was supported by grants from the National Board of Education in Sweden and from the Tri-Centennial Fund of the Bank of Sweden. The director of the project was Karl-Georg Ahlström, Assistant Professor. The pre-run series was undertaken as field-work for an Honours thesis be B. Samuelsson and A. Svärdström, who also did the testing in the main part of the project.

REFERENCES

Bjurggren, G. (1954), A method to test the intelligibility of the speech of pre-school children with severe hearing impairment. Acta Oto-Laryngologica. Suppl, 110. Stockholm 1954.

Björk, L. (1961), Velopharyngeal function in connected speech. Studies using tomography and cineradiography synchronized with speech spectrography. Acta Radiol. Supl. 202. Stockholm.

Fant, G. (1958), Modern instruments and methods for acoustic studies of speech. Proc. 8th Int. Congr. Ling., Oslo, 1958.

Fant, G. (1962), Sound spectrography. Proc. 4th Int. Congr., Phonetic Sciences, Helsinki 1961. Mouton & Co 1962.

Guilford, J.P. (1965), Fundamental stitistics in psychology and education. Mc Graw-Hill, Tokyo. International student edition, Fourth edition.

Hadding-Koch, K. & Pettersson, L. (1964), Instrumentell fonetik. En handledning. Gleruns Lund.

Hansson, G. (1965), Hur mycket förstår hörande av döva elevers tal efter 8-års skolundervisning? Sveriges dövas riksförbunds forskningsserie nr 5.

Hansson, G. (1967), Dimensions in speech sound perception. An experimental study of vowel perception. Inaugural dissertation. Ericsson Technics, No 1.

Hudgins, C.V. & Numbers, F. C. (1942), An investiagtion of the intelligibility of the speech of the deaf. Genet. Psychol. Monog., 25, 289 -332.

Mac Neilage, F. F., Rootes, T.P. & Chase, R.A. (1967), Speech production in a patient with severe impairement of Somesthec percention and motor control. J. Speech Hearing Res., 10, 447 - 449.

Martony, J. (1966), Röstlägeskorrelation. Nord. Tidskr. f. Dövunderv. 3, 102 - 106.

Martony, J. (1967), On the correlation of the voice pitch level for souerely hard of hearing subjects. Speech read at the "Conference on speech analyzing aids for the deaf", Gallaudet College, Washington D.C. 1967.

O'Neill, J. J. (1954), Contributions of the visual components of oral symbols to speech comprehension. J. Speech Hearing Dis. 19, 429 - 439.

Phillips, N. D., Remillard, W., Pronovost, W. & Bass, S. (1967), Teaching of intonation to the deaf by visual pattern matching. Speech read at the "Conference on speech analyzing aids for the deaf", Gallaudet College, Washington D. C. 1967.

Prather, E.M. (1960), Scaling defectiviness of articulation by direct magnitud-estimation. J. Speech Hearing Res. 3, 380 - 392.

Pronovost, W., Yankin, U., Andersson, D. C. & Lerner, R. (1967), The voice visualizar. Speech read at the "Conference on speech analyzing aids for the deaf", Gallaudet College, Washington D.C., 1967.

Risberg, A. (1966), Talanalys och talträning. Nord. Tidskr. f. Dövunderv. nr. 2, 57 - 69.

Risberg, A. (1967), Synpunkter på den "kybernetiska metoden". Nord. Tidskr. f. Dövunderv. nr 4, 90 - 100.

Stark, R.E. (1967), Preliminary work with the New Bell Telephone Visible Speech Translator. Speech read at the "Conference on speech analyzing aids for the deaf", Gallaudet College, Washington D.C. 1967.

Sherman & Moodie, C.E. (1957), Four psychologocal scaling methods applied to articulation defectiveness. J. Speech Hearing Dis. 23, 698 - 706.

Steward, R. B. (1967), "By ear alone". Speech read at the "Conference on speech analyzing aids for the deaf". Gallaudet College, Washington D.C. 1967.

Proportions of correct responses (p) in percent and biserial correlations (r_{bis}) for tasks coded according to Version 1, distributed over judges and grades.

Grade 2

	Judge 1		Judge 2		Judge 3		Judge 4		1 - 4
Item No.	p	r_{bis}	p	r_{bis}	p	r_{bis}	p	r_{bis}	p
1	8	.51	11	.48	5	.30	11	.51	9
2	38	.28	32	.41	38	.38	35	.42	36
3	16	.23	24	.33	22	.41	24	.50	22
4	43	.41	51	.45	43	.45	46	.47	46
5	35	.58	32	.41	27	.46	32	.48	32
6	22	.28	32	.30	27	.24	30	.43	28
7	16	.26	8	.18	11	.26	14	.39	12
8	19	.57	19	.38	22	.58	24	.45	21
9	16	.43	24	.38	22	.27	22	.19	21
10	11	.20	16	.23	11	.26	5	.33	11
11	19	.34	14	.33	19	.36	22	.26	18
12	11	.43	19	.39	14	.38	14	.38	14
13	24	.28	19	.41	22	.52	14	.41	20
14	32	.31	32	.44	30	.43	19	.43	28
15	38	.30	38	.39	43	.30	43	.31	41
16	22	.27	30	.38	30	.12	38	.24	30
17	14	.53	14	.54	14	.50	16	.48	14
18	19	.32	22	.28	22	.38	22	.26	21
19	3	.03	8	.41	8	.55	14	.46	8
20	16	.32	19	.34	22	.41	14	.18	18
21	24	.26	22	.24	22	.33	22	.27	22
22	11	.39	22	.52	16	.41	16	.40	16
23	11	.22	11	.27	14	.35	11	.34	11
24	11	.53	19	.54	14	.52	8	.41	13
25	27	.43	24	.46	27	.32	24	.21	26
26	35	.45	32	.36	35	.26	30	.60	33
27	8	.34	14	.22	8	.32	8	.19	9
28	27	.39	22	.51	24	.35	22	.32	24
29	19	.14	19	.27	19	.03	16	.03	18
30	14	.32	14	.24	19	.29	16	.40	16

Proportions of correct responses (p) in percent and biserial correlations (r_{bis}) for tasks coded according to Version 1, distributed over judges and grades.

Grade 4

	Judges									
	1		2		3		4		1 - 4	
Item No.	p	r_{bis}	p	r_{bis}	p	r_{bis}	p	r_{bis}	p	
1	21	.30	23	.56	10	.33	23	.41	22	
2	48	.09	48	.02	50	.03	48	.07	49	
3	29	.00	33	.75	33	.74	20	.59	31	
4	54	.60	58	.68	50	.57	50	.70	54	
5	52	.77	58	.71	46	.71	46	.60	50	
6	46	.51	48	.60	44	.50	40	.51	45	
7	29	.57	31	.54	27	.51	33	.54	30	
8	33	.05	33	.60	46	.09	38	.48	39	
9	21	.45	33	.68	31	.47	27	.35	28	
10	17	.25	25	.60	15	.40	13	.37	18	
11	35	.59	27	.55	33	.66	31	.62	31	
12	42	.66	38	.63	33	.58	44	.63	39	
13	27	.41	27	.43	33	.44	23	.31	27	
14	44	.76	48	.75	40	.58	33	.61	41	
15	35	.59	35	.69	40	.61	35	.50	36	
16	31	.70	31	.67	42	.71	37	.58	35	
17	35	.56	29	.51	35	.64	35	.55	33	
18	29	.38	41	.39	35	.41	27	.39	33	
19	31	.59	38	.57	37	.59	23	.55	30	
20	31	.56	33	.60	33	.53	37	.54	33	
21	29	.42	27	.51	25	.48	1	.40	25	
22	38	.61	50	.73	44	.69	44	.61	44	
23	21	.51	19	.54	25	.53	15	.46	20	
24	37	.55	33	.27	33	.54	29	.50	33	
25	35	.45	37	.44	46	.47	42	.38	40	
26	44	.59	44	.62	48	.70	37	.50	43	
27	37	.60	35	.52	37	.67	33	.54	35	
28	37	.61	40	.60	40	.72	38	.69	39	
29	33	.54	40	.58	29	.47	33	.17	34	
30	25	.34	23	.35	23	.45	25	.52	24	

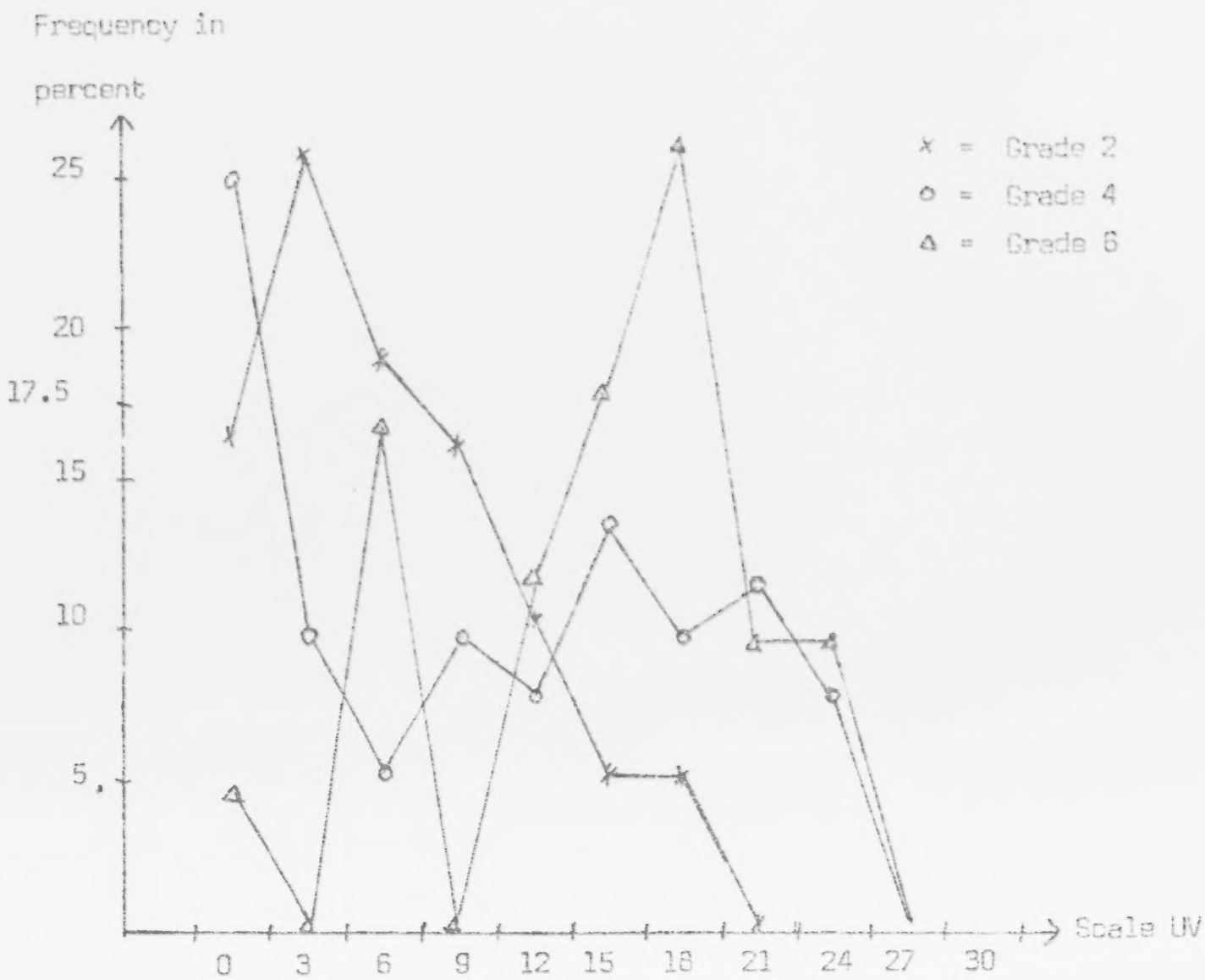
Proportions of correct responses (p) in percent and biserial correlations (r_{bis}) for tasks coded according to Version 1, distributed over judges and grades.

Grade 6

	Judges									
	1		2		3		4		1 - 4	
Item No.	p	r_{bis}	p	r_{bis}	p	r_{bis}	p	r_{bis}	p	
1	32	.30	50	.46	36	.40	27	.53	36	
2	68	.60	59	.64	59	.63	59	.47	61	
3	50	.43	45	.46	45	.55	36	.14	44	
4	73	.08	77	.26	82	.59	73	.19	76	
5	68	.64	68	.56	68	.67	64	.56	67	
6	55	.09	72	.15	64	.55	68	.34	65	
7	32	.56	27	.58	27	.35	32	.57	30.	
8	32	.26	45	.38	45	.43	45	.29	42	
9	32	.07	32	.57	27	.58	27	.54	30	
10	27	.10	36	.21	23	.07	23	.02	27	
11	50	.20	55	.33	50	.35	64	.34	57	
12	50	.22	68	.32	64	.25	50	.13	63	
13	55	.18	50	.26	50	.00	45	.12	52	
14	50	.32	50	.41	36	.45	50	.32	40	
15	50	.64	55	.51	55	.64	59	.60	57	
16	45	.50	50	.51	50	.59	64	.42	52	
17	41	.21	50	.44	36	.40	36	.38	41	
18	36	-.74	41	.08	45	-.06	45	.09	42	
19	32	.47	23	.37	23	.37	32	.31	27	
20	50	.61	45	.71	55	.39	50	.49	52	
21	21	.10	27	.42	36	.19	45	.13	38	
22	36	.37	50	.33	36	.21	45	.43	42	
23	45	.24	32	.42	36	.43	41	.18	38	
24	32	.28	45	.44	45	.70	50	.58	43	
25	55	.26	59	.16	59	.40	73	.48	61	
26	55	.23	45	.75	64	.29	59	.23	56	
27	55	.41	45	.52	55	.05	55	.68	52	
28	55	.43	50	.05	55	.52	59	.37	55	
29	41	.56	36	.38	41	.25	41	.33	40	
30	23	.24	45	.45	32	.36	41	.41	35	

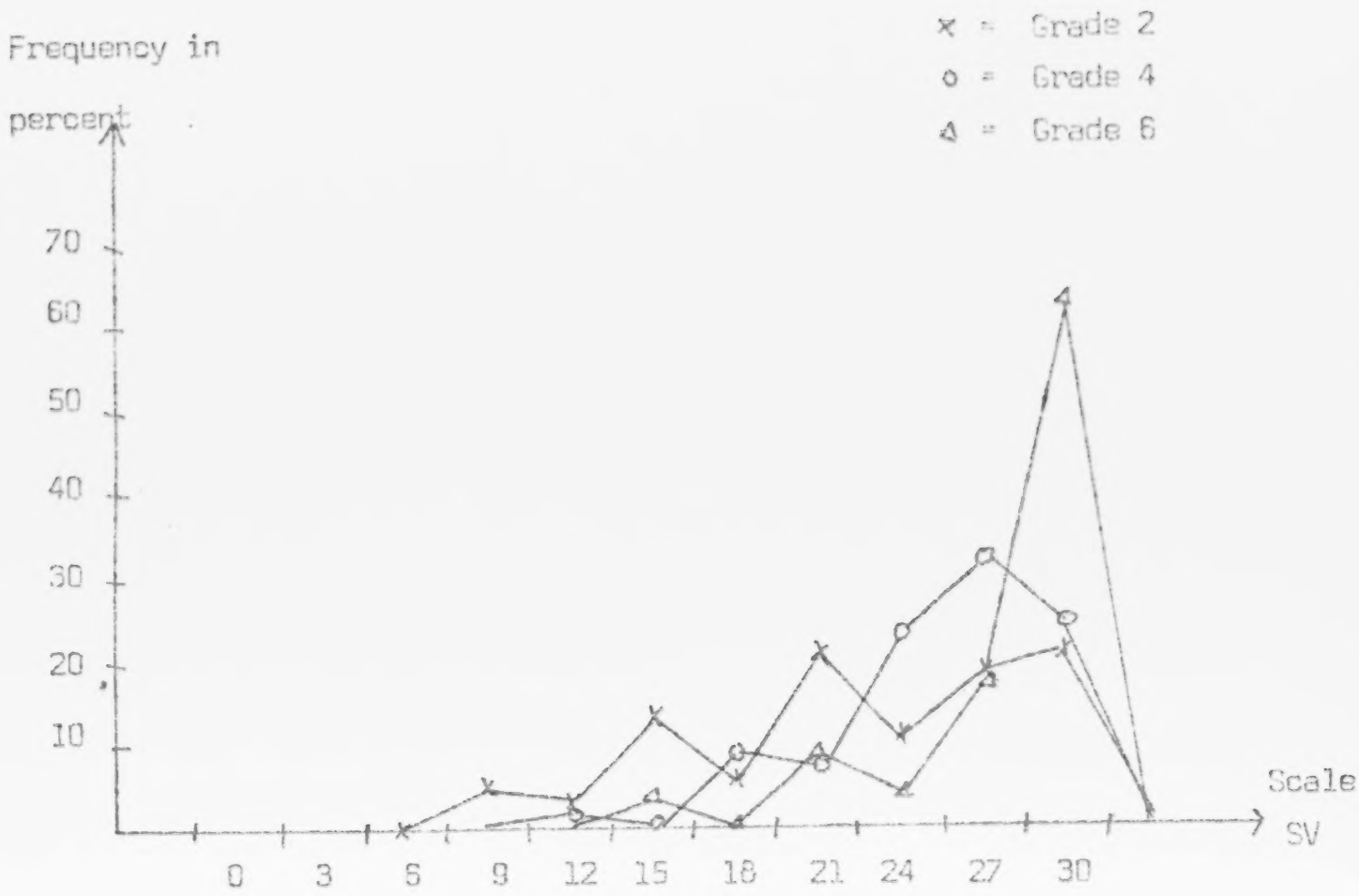
Scale UV

Frequency Polygon for Grades 2, 4 and 6 with frequencies expressed in percentages.



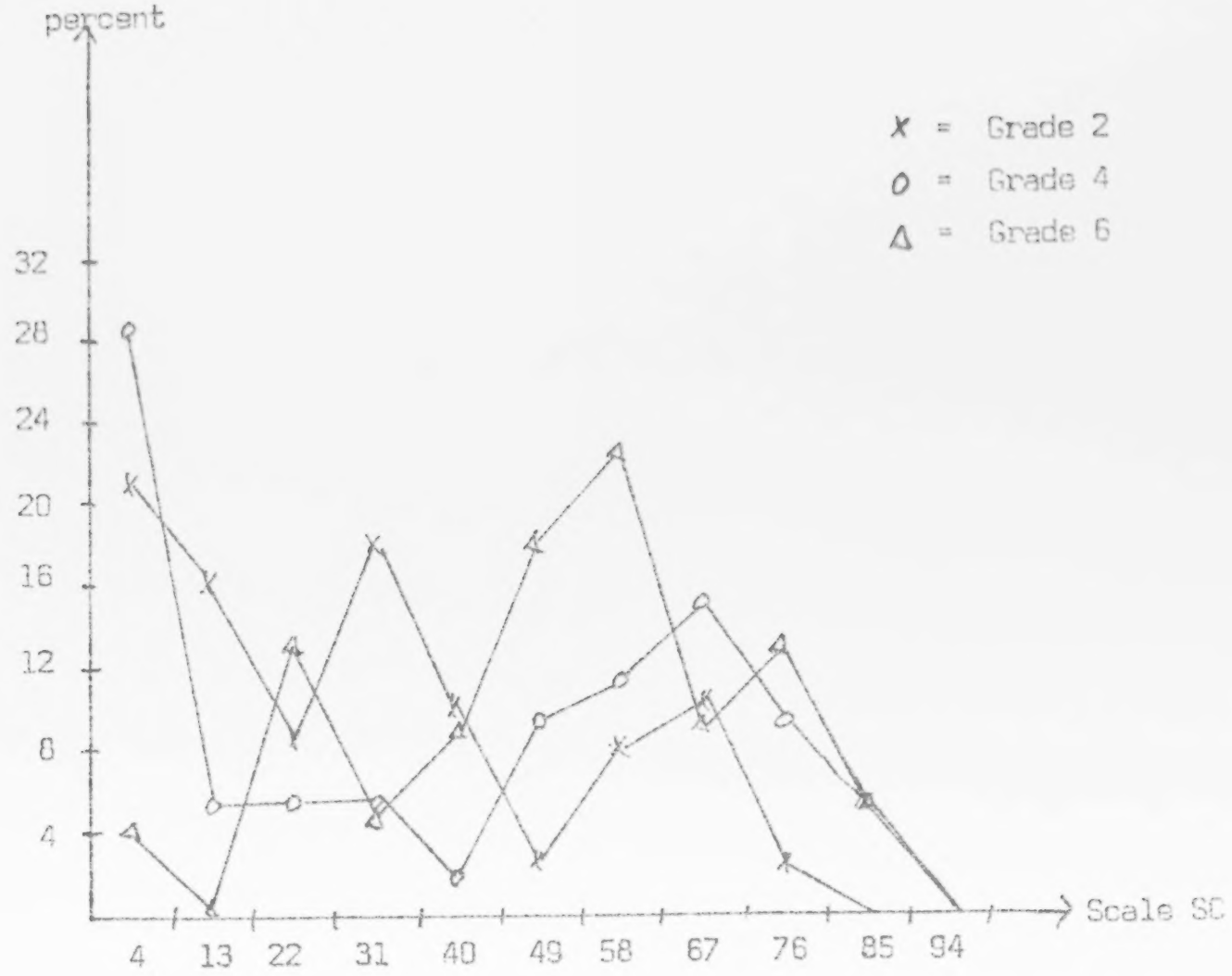
Scale SV

Frequency Polygon for Grades 2, 4 and 6 with frequencies expressed in percentages.



Scale SC

Frequency Polygon for Grades 2, 4 and 6 with frequencies expressed in percentages.



Name of Child:_____ Grade:_____ Series:_____

Word	Doubtful	Certain	Result
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			
21.			
22.			
23.			
24.			
25.			
26.			
27.			
28.			
29.			
30.			

Instructions (to the judges)

We have shown 30 pictures of concrete objects to about 100 children with hearing-defects. They were asked to say out loud what the pictures represented - this was recorded on tape.

Your task is to listen to these tapes, write down on your provided answer-sheets every word you hear and estimate its comprehension along two dimensions in which 1 means Doubtful and 2 means Understood with Certainly. If you do not understand the word, mark a dash. If either the child or we have said "Don´t know", mark an x.

At the top of the answer-sheet, please write the name of the child and the number of the series; after your last judgement, write your own name. You are to work independantly of each other.

Finally we would like to draw your attention to the fact that you should completely disregard any comments we have made during the taping.

Are there any questions?